Journal of Theoretical and Applied Information Technology
15th September 2022. Vol.100. No 17

ISSN: **1992-8645** www.jatit.org E-ISSN: **1817-3195**

GENERIC FRAMEWORK FOR EARLY PREDICTION OF HEART DISEASE USING LEARNING METHOD

VENKATESWARARAOCHEEKATI[1] , DR .D.NATARAJASIVAN [2] , DR.S.INDRANEEL [3]

[1] Research Scholar, Department .of CSE Annamalai University,T.N,
[2] Assistant.Prof., Department .ofCS,AnnamalaiUniveristy
[3] Professor,Dept.of CSE ,.Ann's Engineering College, Chirala

E-mail: chvraograce@gmail.com[1], natarajasivan@gmail.com[2], sreeram.indraneel@gmail.com[3]

ABSTRACT

The incidence of heart disease is growing at a startling rate, making identification at an earlier stage all the more important. The identification of heart illness is a challenging endeavour that calls for accuracy and dexterity on the part of the diagnostician. To determine which people, based on a variety of different medical criteria, are more likely to develop heart disease, the purpose of this study is to identify those patients. Models for determining the likelihood of a person being diagnosed with a cardiac disease have been developed, and these models can be used to evaluate an individual's risk.SGD Classifier, Additional Tree Classifier, Calibrated ClassifierCV, Gaussian Mixture, Nearest Centroid, MultinomialNB, Logistic RegressionCV, Linear SVC, Linear Discriminant Analysis, SGD Classifier, Calibrated ClassifierCV, Linear SVC, Linear Discriminant Analysis, SGD Classifier, Calibrated ClassifierCV, Quadratic Discriminant Analysis, GaussianNB, Random Forest Classifier, ComplementNB, MLP Class; Gaussian Mixture; Nearest Centroid;, LGBM Classifier, Ada Boost Classifier, A very helpful method was utilised in order to determine how the model might be applied to improve the level of accuracy with which heart disease can be predicted in any given individual. Using Deep Learning and Random Forest Classifiers, which had a high degree of accuracy compared to previous models, the suggested model was able to effectively predict a person's likelihood of acquiring heart disease. The algorithm for predicting cardiac disease that has been proposed would not only make medical care more effective but would also reduce expenses. By making use of this knowledge, we are able to improve our ability to anticipate who might develop heart disease in the future. The model was constructed in Python, and the data was taken from the Kaggle vault.

Keywords: *Disease of the Heart, Artificial Intelligence, Deep Learning, Prediction*

1. INTRODUCTION

Machine learning, also known as ML, is a subfield of artificial intelligence (AI) that enables computers to acquire knowledge and improve themselves as a result of experience without being specifically programmed to do so.Artificial Intelligence (AI) is a subset of ML. The purpose of machine learning is to develop computer models that are capable of acquiring knowledge independently through interaction with data [11-20]. Students can be assisted in seeing patterns in data and improving their ability to make decisions in the future by applying what they have learned through the use of examples, first-hand experience, or instructions [21-30]. The primary objective is to give computers the ability to learn and behave like people while also enhancing their learning on their own over time. This is accomplished mostly through the utilisation of data and information taken from the actual world [31-40].

Deep learning is a subfield of Artificial Intelligence (AI) that attempts to model the method in which people acquire specific categories of information [41-50]. The field of data science also includes important subfields such as statistical analysis and predictive modelling. Data scientists are now able to collect, analyse, and interpret massive amounts of data in a more expedient and straightforward manner because to deep learning.

Cardiovascular diseases encompass a broad category of ailments, each of which has the potential to have an adverse effect on a person's heart (CVD). According to the Centers for Disease Control and Prevention (CDC), coronary heart disease is the leading cause of death across all racial and ethnic groups in the United States (African Americans, American Indians and Alaska Natives, and white people). In the United States, the three most frequent risk factors for developing

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

cardiovascular disease are hypertension (high blood pressure), high cholesterol, and smoking. Diabetes, obesity (high BMI), inactivity (lack of physical exercise), and excessive alcohol use are other key markers, as illustrated in Table 1. In the field of healthcare, it is of the utmost importance to locate and implement preventative measures for the heart disease risk factors that have the largest influence. After then, algorithms for deep learning and machine learning can be applied to the data in order to find "patterns" in the information that can be used to determine a patient's current state of health [71-80].

For the purposes of this dataset, the Kaggle depository is searched for information regarding the patient's medical history and other relevant details. By analysing this dataset, we will be able to establish whether or not a patient has cardiac disease. Within the dataset, a total of 18 different medical traits, qualities, and variables are included...

2. RELATED WORK

A significant amount of previous research into the use of Machine Learning strategies to the diagnosis of cardiovascular heart disease served as the impetus for this investigation. It is possible to make an accurate prognosis of cardiovascular heart disease by making use of a number of different methodologies, some of which are given below.

There is the possibility of incorporating regression models such as Logistic Regression, KNN, and Random Forest Classifier. The findings make it very evident that each tactic possesses a unique capacity to accomplish the objectives that were established.

MukteviSrivenkatesh did research on the prediction of cardiovascular illness by analysing the accuracy of applying rules to a dataset that was collected in a region and comparing it to the individual findings of a Support Vector Machine, Random Forest, Naive Bayes classifier, and logistic regression. This was done so that he could better understand how accurate the application of rules can be in predicting the occurrence of cardiovascular illness. The machine learning algorithms that were utilised in this study had an accuracy that ranged from 58.71 percent to 77.06 percent when it came to determining whether or not a patient had cardiovascular disease. When compared to other machine learning algorithms, logistic regression had the highest accuracy (77.06 percent), making it the most effective of the available machine learning techniques.).

The authors of [82] presented a model that was based on supervised learning approaches such as Nave Bayes, decision trees, K-nearest neighbours, and the random forest algorithm. This model incorporated a variety of factors that are associated with heart disease. The Cleveland database at UCI, which consists of people with cardiac disease, served as a source of data for this study. 303 occurrences and 76 attributes are contained within the data collection. We only examined 14 of the 76 available features because these particular qualities are essential to demonstrating the effectiveness of different algorithms. According to the findings, the K-nearest neighbour algorithm has the best accuracy rating, clocking in at 90.79 percent.

The UCI Machine Learning Heart Disease dataset was utilised in order to evaluate and contrast the results of [83] various machine learning approaches as well as deep learning. In order to carry out the study, the dataset consists of fourteen primary attributes. Accuracy and confusion matrices are utilised in the verification process for a variety of potential results. The findings are improved by using Isolation Forest in order to eliminate some aspects from the dataset that are deemed to be unnecessary, and the data itself are also standardised. Through the application of a technique known as deep learning, an accuracy rate of 94.2 percent was accomplished.

During the course of the research, they utilised a total of nine different machine learning classifiers, which included AB, LR, ET, MNB, CART, SVM, LDA, RF, and XGB [84]. They did this by preprocessing the data, normalising the data, and tweaking the hyperparameters, and then testing their results on a typical dataset for heart disease. They not only validated and trained the machine learning algorithms, but also did it with the assistance of K-fold cross validation. It was discovered that the accuracy of prediction classifiers improved with hyperparameter tuning, and remarkable results were obtained with data standardisation and the hyperparameter tuning of machine learning classifiers in the experimental results. [Citation needed] [Citation needed] [Citation needed]

In the publication, [85] the authors propose using methods from machine learning to make predictions about cardiovascular disease. In making their projection, they considered the BMI to be one of the most critical elements. It was discovered that using body mass index (BMI) as a predictor of cardiovascular disease was successful. The major purpose of the study was to investigate the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

relationship between body mass index and the danger of developing cardiovascular disease. Several distinct aspects of the model, in addition to regression and classification methods, have been proposed as possible additions to the model. The researchers found that body mass index (BMI) was an important indicator of cardiovascular disease.

In the research conducted by [86], evolutionary techniques such as the Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) were employed to further improve the accuracy of machine learning systems. For the purpose of feature selection, we made use of genetic algorithms and particle swarm optimization in conjunction with the Nave Bayes (NB), Support Vector Machine (SVM), and J48 algorithms. The performance of the feature selection process is evaluated first, before machine learning is applied to either the whole dataset or the reduced dataset. Several different machine learning algorithms, such as NB, SVM, Decision Tree (DT), Logistic Regression (LR), and Random Forest, have been put to the test in an effort to forecast the occurrence of heart illness (RF). The findings indicate that the GA is the most effective algorithm for feature selection due to the fact that it results in the largest improvement in prediction accuracy.

After reviewing the previous research, it has become clear that the dataset that was used was the UCI Machine Learning Heart Disease dataset [87], which is distinct from the dataset that was stored in the Kaggle repository. These algorithms will be trained, validated, and tested with the assistance of other machine learning algorithms in addition to a revolutionary deep learning technique...

3. DATASET

Every individual who was included in the Organized Dataset underwent stringent screening based on their family history of cardiovascular disease as well as other medical issues. Diseases that can affect the heart are grouped together under the umbrella term "heart disease." According to the World Health Organization (WHO), cardiovascular problems are the major cause of death among people in the middle years of their lives. The majority of the information acquired came from the medical histories of 319795 people who ranged in age from infants to the elderly. This dataset provides all of the information that we require to make an educated judgement regarding whether or not a patient has a heart disease, so we can use it to make that determination (as seen in Table 1). This dataset contains a total of 304 patients and 18 medical characteristics that can assist us in determining whether or not a patient is at risk of acquiring a cardiac ailment. The total number of patients included in this dataset was 304. The data were separated into these three categories: training, validation, and testing. The entirety of this dataset, which consists of 319795 rows and 18 columns, is a single record [82]. As can be seen in Table 1, all of the traits that are pertinent to this discussion have been included...

4. METHODOLOGY

Throughout the course of this research, a number of different machine learning strategies—including Gaussian Mixture and Nearest Centroid, MultinomialNB, and Logistic RegressionCV—have been utilised.are investigated. In this particular piece of research, the methods GaussianNB and Random ForestClassifer were utilised. ComplementNB, MLP Classifier ComplementNB, Bernoulli Classifier Bernoulli Classifier Classifiers based on the LGBM and the Ada Boost K Nei Classifier K Nei Classifier This investigation takes the form of a literature survey, during which published works on cardiovascular illness as well as already-existing databases are analysed. The method consists of a series of methods that transform raw data from a dataset into usable information for a variety of purposes, the amount of skill possessed by the customers. The proposed method, which can be seen illustrated in Figure 1, is broken down into the following stages: The process of preprocessing consists of three stages: the first is the collecting of data, the second is the extraction of significant values, and the third is the investigation of the collected data. Preprocessing includes identifying and filling in missing data values, as well as cleaning and normalising the data [81]. A crucial part of the classification process involves applying a classifier to the previously preprocessed data in order to organise it. In the final stage of testing, we used a variety of performance metrics to evaluate our model and determine whether or not it was reliable and productive. An accurate method for predicting the risk of heart disease has been developed here by making use of a number of different classifiers. When making its projections, this model takes into account a wide range of factors, including age, body mass index (BMI), frequency of smoking and drinking, risk of stroke, state of physical and mental health, and DiffWalking. [82]..

Journal of Theoretical and Applied Information Technology
15th September 2022. Vol.100. No 17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table1: Shows The Features Of The Dataset

Feature	Description	Type of Feature
Heart Disease	Respondents that have ever reported having coronary heart disease (CHD) or myocardial infarction (MI).	Yes/No, Goal
BMI	Body Mass Index (BMI).	Yes/No
Smoking	Have you smoked at least 100 cigarettes in your entire life?	Yes/No
Alcohol Drinking	Heavy drinkers (adult men having more than 14 drinks per week and adult women having more than 7 drinks per week	Yes/No
Stroke	(Ever told) (You had) a stroke?	Yes/No
Physical Health	Now thinking about your physical health, which includes physical illness and injury, for how many days during the past 30 days was your physical health not good? (0-30 days).	Numeric (0-30 days)
Mental Health	Thinking about your mental health, for how many days during the past 30 days was your mental health not good? (0-30 days).	Numeric (0-30 days)
Diff Walking	Do you have serious difficulty walking or climbing stairs?	Yes/No
Sex	Are you male or female?	Male/Female
Age Category	Age category.	Category (14 groups)
Race	Imputed race/ethnicity value.	Category
Diabetic	(Ever told) (You had) diabetes?	Yes/No
Physical Activity	Adults who reported doing physical activity or exercise during the past 30 days other than their regular job.	Yes/No
Gen Health	Would you say that in general your health is...	Category
Sleep Time	On average, how many hours of sleep do you get in a 24-hour period?	Numeric
Asthma	(Ever told) (You had) asthma?	Yes/No
Kidney Disease	Not including kidney stones, bladder infection or incontinence, were you ever told you had kidney disease?	Yes/No
Skin Cancer	(Ever told) (You had) skin cancer?	Yes/No

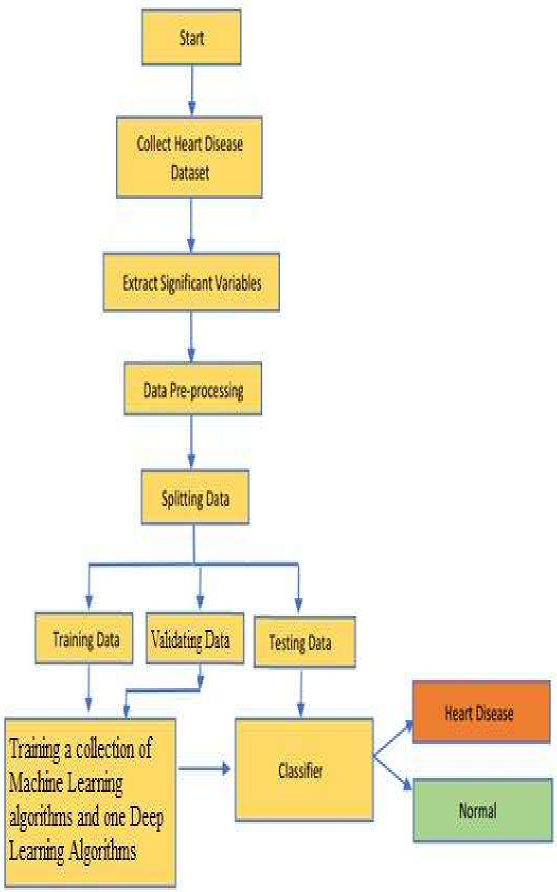

Figure 1: Proposed System Model

4.1 Data Cleaning

There are approximately 319795 records in the dataset, and a total of 18 different columns are available for selection. There are four qualitative characteristics along with fourteen others that are quantitative. It is possible to convert string properties that have only two possible values; however, before we do so, we need to make sure that there aren't any unexpected ones in the mix. In the preprocessing stage, we will make use of OneHotEncoder due to the fact that certain characteristics can take on more than two distinct values...

4.2 Exploratory Analysis

4.2.1 Analyzing and Displaying Feature Categories
Figure 2 depicts an individual's sexual orientation as a factor in determining whether or not they have heart disease. The x-axis is labelled with the numbers 1 and 0, which stand for males and females, respectively. The following can be deduced from the diagram:

In general, males are at a greater risk of acquiring cardiovascular disease than women are**.**

• The risk of coronary disease in women is lower than it is in men on average.

In Figure 3, we have a representation of the prevalence of heart illness in connection to the patient's smoking status. The following is a conclusion that may be drawn from the graph**:**
Even non-smokers have been discovered to have heart disease, which suggests that the condition is caused by something other than smoking.

Figure 4 shows that there is a significant disparity between the percentage of people of different races who have cardiac disease and the percentage of people of different races who do not have cardiac illness. From the diagram, we are able to deduce the following:

The incidence of cardiovascular disease is significantly higher among people of colour.

Figure 5 presents a comparison of patients who had cardiac disease with those who did not have the condition, broken down by age group. As shown in Figure 5, those who have reached the age of 80 or older have an increased likelihood of having heart disease.

Figure 6 depicts the distribution of occurrences with heart disease based on the KidneyDisease feature. According to the graph, individuals who do not have any form of kidney disease are at a greater risk of developing heart disease.

Figure 7 depicts cases of cardiovascular illness that may or may not be accompanied with skin cancer. The following can be inferred from the diagram:

• Those who have not been diagnosed with skin cancer are less likely to suffer from heart disease.

The distribution of cases with yes/no heart disease based on previous exposure to stroke risk factors is shown in Figure 8. The inferences that can be drawn from the diagram are as follows.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

• Those who have previously suffered from a stroke have a reduced risk of developing heart disease in the future.
• Cases of heart illness that were caused by previous exposure to diabetic conditions are shown in figure 9. The graphic demonstrates the following: • People who have a family history of diabetes have a lower risk of developing heart disease.

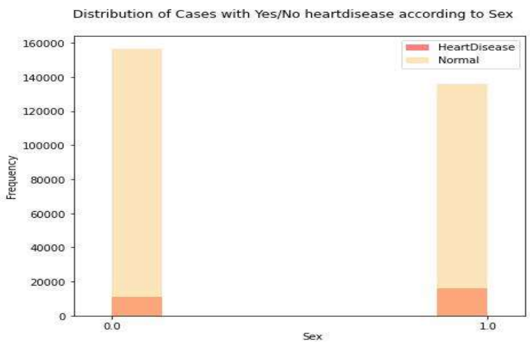

Figure 2: The Sex-Based Distribution Of People Who Have Or Don't Have Heart Disease

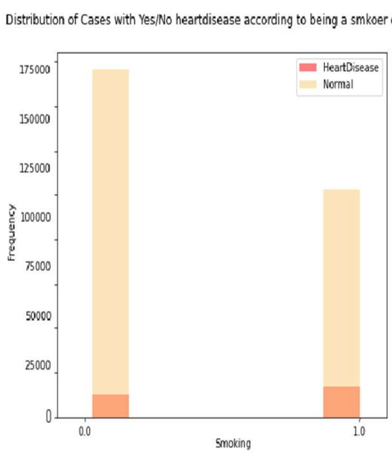

Figure 3According To Whether Or Not A Person Smokes, Below Is The Breakdown Of People Who Have Heart Disease:

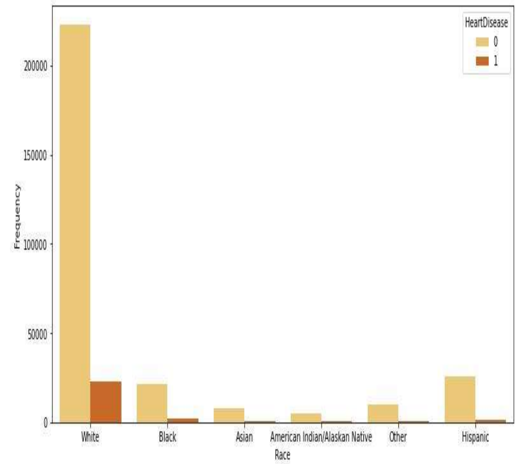

Figure 4 By Race, The Percentage Of Those With

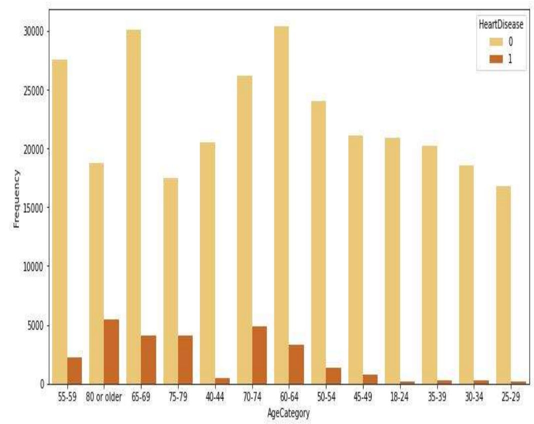

Figure 5: Cases With Or Without Heart Disease, As Indicated By The Agecategory Heart Disease And Those Without It.

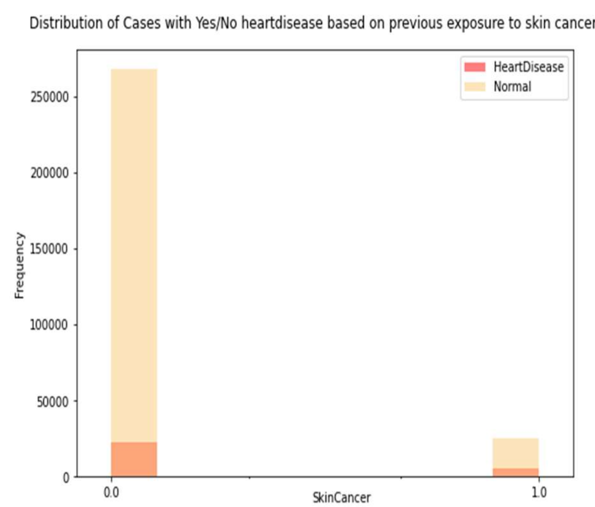

Figure 6: According To Kidneydisease, The Incidence Of Heart Disease Is Distributed As Follows

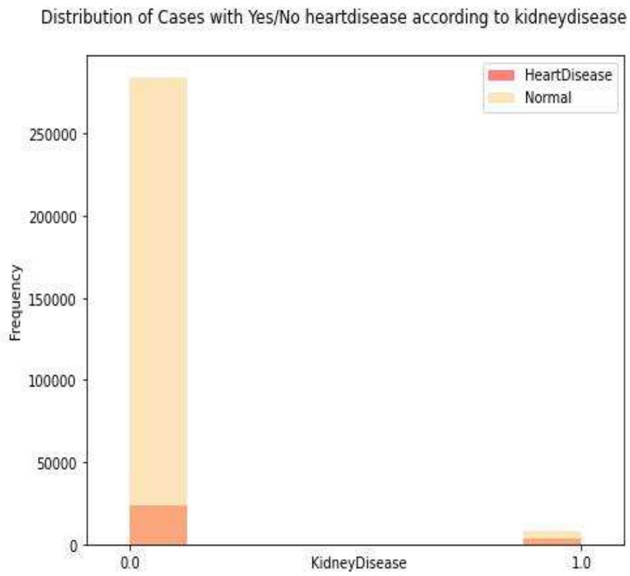

Figure 7: No Heart Disease Versus Heart Disease Cases According To Skincancer

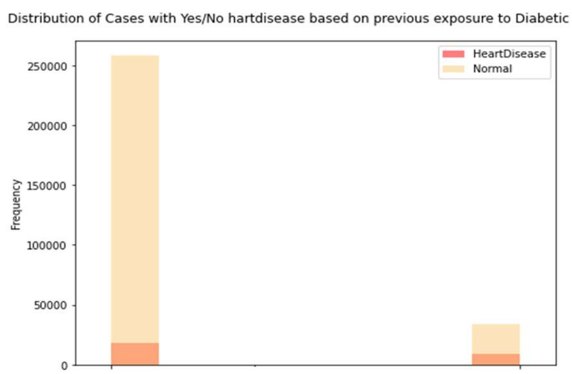

Figure 8: Based On Previous Exposure To Stroke, The Distribution Of Cases With Heart Disease And No Heart Disease

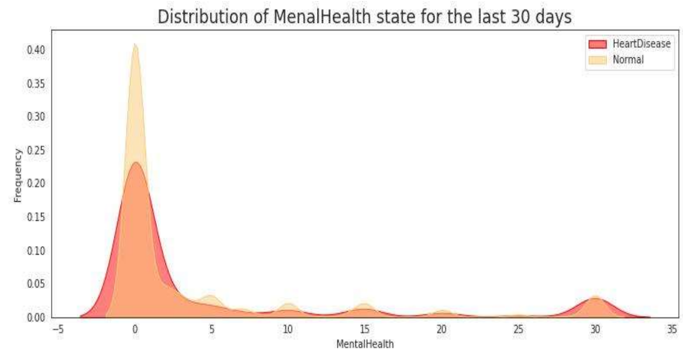

Figure 9: If A Patient Had Previously Been Exposed To Diabetic, The Following Is The Likelihood That They Have Heart Disease

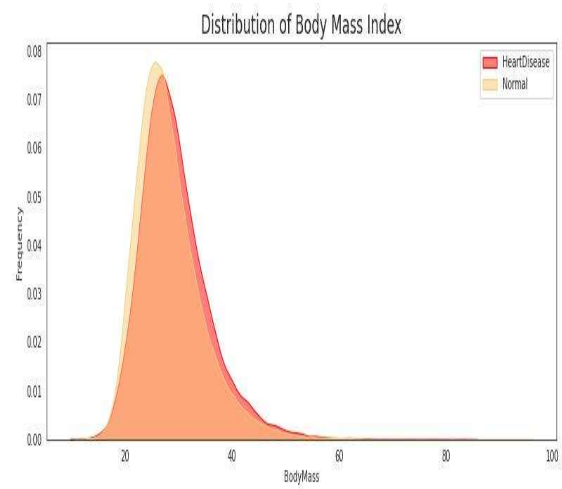

Figure 10: Cases Classified As Having Or Not Having A Heart Condition, As Determined By Their Body Mass Index

4.2.2 Visualization of Numerical Features

Figure 10 makes it very evident that individuals who weigh less than 40 kg are at an increased risk of developing cardiovascular disease. Figure 11 shows a representation, derived from the data provided by SleepTime, showing the distribution of individuals who have and do not have heart diseaseThe number of people who have heart disease is depicted in figure 12, which is based on the individuals' current physical health status and how it has changed over the course of the last 30 days. The number of cases of heart disease that can be attributed to a person's mental

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

health status over the past thirty days is depicted in Figure 13, which can be found here...

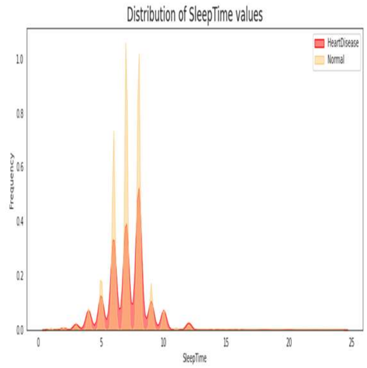

Figure 11: Based On Sleeptime Measurements, The Distribution Of Patients With And Without Cardiac Disease Was Calculated.

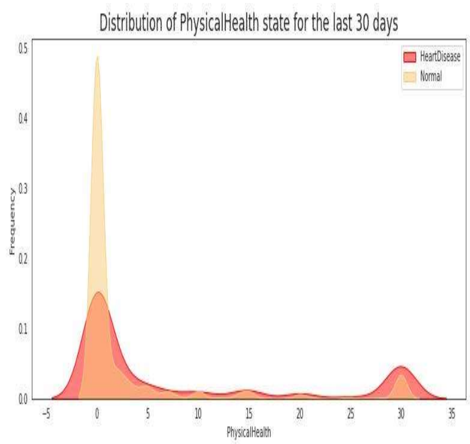

Figure 12Based On The Last 30 Days Of Physical Health, The Percentage Of Patients With Heart Disease Or No Heart Disease

.

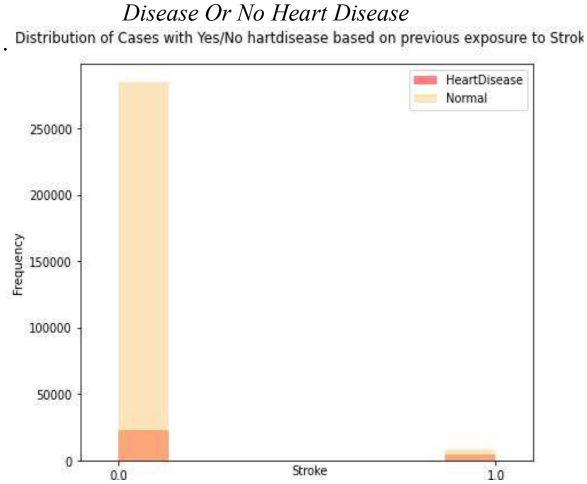

Figure 13: Based On The Last 30 Days Of Mental Health, The Percentage Of People Had Heart Disease Or No Heart Disease

4.3 Data Pre-processing

4.3.1 Standardization

The following characteristics were standardised with the use of the StandardScale function: MentalHealth, BMI, PhysicalHealth, and SleepTime.

4.3.2 Encoding

The following characteristics were encoded with the One Hot Encoding method: age category, race, and general health.

4.4 Bringing Harmony to the Dataset

There are 91 times more cases of heart disease in the non-target category, which is heart disease, than there are in the target category, which is heart disease. (You're right, it's heart disease). Because of this, the dataset does not have a balanced distribution of values..

Over sampling and under sampling are two common strategies that are utilised while attempting to achieve a balanced dataset. We use the under sampling method when we want to decrease the value of the higher attribute so that it is equivalent to the value of the lower attribute, In addition, when we need to raise the value of a lower attribute, we make use of a technique called "over sampling." so that it is equivalent to the value of the higher attribute.

We considered both under and over sampling to be a part of the parameters of this investigation, and we compared the results obtained from each of these two approaches...

4.5 SSplitting Dataset

After all of the preparation was finished, the dataset was split up into three separate datasets: one for training, one for validating, and one for testing. The breakdown of the divide is as follows: sixty percent, twenty percent, and twenty percent.

4.6 First Experiment

In the first experiment that was carried out, under sampling was used as the method for achieving a more equitable distribution of values across the dataset. We trained the dataset using a

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

variety of machine learning methods, one of which was a deep learning algorithm that was built from scratch by our team. Metrics such as accuracy, precision, recall, f1-score, and the total amount of time needed for training were utilised in order to keep track of the results of the training. The results of each algorithm are detailed in Table 2, which can be viewed by clicking this link..

Table 2: The Outcomes Of Every Algorithm That Employed The Under Sampling Method

Model Name	Accuracy	Precision	Recall	F1_score	Time in Second
GaussianMixture	0.48817	0.00000	1.00000	0.00000	0.49
Perceptron	0.70283	0.70956	0.69045	0.69887	0.17
NearestCentroid	0.72886	0.74784	0.69354	0.71967	0.12
MultinomialNB	0.74575	0.75284	0.73449	0.74355	0.10
LogisticRegressionCV	0.76320	0.75429	0.78326	0.76850	7.99
LinearSVC	0.76274	0.75528	0.76399	0.76853	1.48
LinearDiscriminantAnalysis	0.76283	0.75532	0.78417	0.76844	0.34
LabelPropagation	0.71178	0.71856	0.69973	0.70902	58.03
CalibratedClassifierCV	0.76265	0.75457	0.78107	0.75739	8.267
ExtraTreeClassifier	0.68100	0.68220	0.68207	0.68214	0.15
SGDClassifier	0.75297	0.73956	0.78417	0.76111	0.17
QuadraticDiscriminantAnalysis	0.75160	0.68905	0.84768	0.76018	0.58
GaussianNB	0.74941	0.70545	0.85951	0.77490	0.14
RandomForestClassifier	0.74420	0.74476	0.74429	0.74414	5.93
ComplementNB	0.75597	0.76469	0.75631	0.75023	0.09
MLPClassifier	0.76110	0.73594	0.81711	0.77440	57.11
BernoulliNB	0.75288	0.76699	0.72903	0.74733	0.15
BaggingClassifier	0.72119	0.73186	0.70136	0.71629	1.79
LGBMClassifier	0.76094	0.75004	0.80328	0.77575	0.68
AdaBoostClassifier	0.76639	0.76588	0.77361	0.76872	1.78
KNeighborsClassifier	0.75662	0.75812	0.74067	0.75839	7.83
LogisticRegression	0.76813	0.76159	0.78308	0.77218	0.82
GradientBoostingClassifier	0.76594	0.75528	0.79345	0.77284	4.61
Deep Learning	**0.78767**	**0.76828**	**0.82602**	0.79611	81.00

Experiment -2

In the second experiment, over sampling was done in order to guarantee that the dataset would be representative of the entire population. We trained the dataset using the exact same collection of machine learning techniques and the exact same deep learning method which was developed from the ground up in exactly the same way. The following metrics were utilised in the recording of the results: accuracy, precision, recall, f1-score, and the total amount of time required for training: Table 3, which can be viewed right here, contains a compilation of the results obtained using each algorithm.

Table 3: The Outcomes Of Every Algorithm That Employed The Under Sampling Method

Model Name	Accuracy	Precision	Recall	F1_score	Time in Second
GaussianMixture	0.48817	0.00000	1.00000	0.00000	0.49
Perceptron	0.70283	0.70956	0.69045	0.69887	0.17
NearestCentroid	0.72886	0.74784	0.69354	0.71967	0.12
MultinomialNB	0.74575	0.75284	0.73449	0.74355	0.10
LogisticRegressionCV	0.76320	0.75429	0.78326	0.76850	7.99
LinearSVC	0.76274	0.75528	0.76399	0.76853	1.48
LinearDiscriminantAnalysis	0.76283	0.75532	0.78417	0.76844	0.34
LabelPropagation	0.71178	0.71856	0.69973	0.70902	58.03
CalibratedClassifierCV	0.76265	0.75457	0.78107	0.75739	8.267
ExtraTreeClassifier	0.68100	0.68220	0.68207	0.68214	0.15
SGDClassifier	0.75297	0.73956	0.78417	0.76111	0.17
QuadraticDiscriminantAnalysis	0.75160	0.68905	0.84768	0.76018	0.58
GaussianNB	0.74941	0.70545	0.85951	0.77490	0.14
RandomForestClassifier	0.74420	0.74476	0.74429	0.74414	5.93
ComplementNB	0.75597	0.76469	0.75631	0.75023	0.09
MLPClassifier	0.76110	0.73594	0.81711	0.77440	57.11
BernoulliNB	0.75288	0.76699	0.72903	0.74733	0.15
BaggingClassifier	0.72119	0.73186	0.70136	0.71629	1.79
LGBMClassifier	0.76094	0.75004	0.80328	0.77575	0.68
AdaBoostClassifier	0.76639	0.76588	0.77361	0.76872	1.78
KNeighborsClassifier	0.75662	0.75812	0.74067	0.75839	7.83
LogisticRegression	0.76813	0.76159	0.78308	0.77218	0.82
GradientBoostingClassifier	0.76594	0.75528	0.79345	0.77284	4.61
Deep Learning	**0.78767**	**0.76828**	**0.82602**	0.79611	81.00

5. RESULTS & DISCUSSIONS

The vast majority of researchers use a variety of algorithms to identify patients who have heart disease, including SVC, Decision tree, KNN, Random Forest Classifier, and Logistic regression. However, as we can see from these results, the highest accuracy of 88.5% is achieved with a dataset that is different from the one that we are using in this study.

In comparison, the algorithms that they used were used by earlier researchers, which means that our algorithms are more accurate, more cost-effective, and faster than the methods that were used by earlier researchers. To put the icing on the cake, Deep Learning was able to achieve the highest possible levels of accuracy with the least amount of time spent training and validating the model. The following are the percentages of accuracy, precision, and recall that the model achieved: accuracy (78.77 percent), precision (76.63 percent), and recall (82.60 percent) (79.51 percent). (81 second time frame). Table 2 displays the remaining results obtained through the application of the various algorithms.

When the same algorithms are applied in conjunction with a greater number of samples, a more accurate version of the same method is generated (Deep Learning). The F1-score was 92 percent, the recall was 94 percent, the accuracy was 92.35 percent with over sampling, the precision was 90.84 percent with over sampling, and the amount of time needed for training and validation was 92.49 percent. Accuracy was found to be 92.35 percent with over sampling, and precision was found to be 90.84 percent. (one minute and one minute and a half).

The Random Forest Classifier came in at a very respectable close second place with a score of 90.21 percent, which it accomplished across the board in terms of accuracy (90.21 percent), precision (90.22 percent), recall (90.21 percent), and F1-score (90.21 percent) (85.51 seconds). The results that were obtained by using the algorithms even after they had been exhausted are presented in Table 3.

We were able to achieve greater precision in our measurements as a direct result of this, in comparison to previous investigations that were carried out in the past. It was discovered that the most successful methods for predicting people who had heart disease were deep learning and the Random Forest Classifier. [Citation needed] [Citation needed] [Further citation is required] [Further citation is required] This demonstrates that Deep Learning and the Random Forest Classifier are superior methods for the earlier diagnosis of heart disease in humans. [Citation needed] [Citation needed,

6. CONCLUSION

Patients' medical histories, including BMI, smoking and alcohol consumption, stroke, physical and mental health, and more, were extracted from a dataset to predict who will get cardiovascular disease and go on to die from it. Deep learning was used to construct a cardiovascular disease diagnosis model that combines a number of different machine learning classification modelling techniques along with one. Patients who have a medical history that includes a diagnosis of heart disease can benefit from having this Heart Ailment detecting device installed in their bodies.

The development of this model made use of a wide variety of different classifiers, including Deep Learning and the Random Forest Classifier. Our models have an accuracy that is 92.23 percent accurate overall. The accuracy of the model's ability to determine if a particular person has a heart condition is improved when the dataset is balanced. [Case in point:] [Case in point:] Because we have used these tactics, we are now in a position to more accurately forecast the condition of the patient, which allows us to drastically cut down on costs. Because machine learning strategies are superior to those of humans and are able to forecast outcomes more accurately, they may be used to a wide variety of medical datasets, which is beneficial for both patients and medical professionals. In conclusion, we were able to acquire a prediction accuracy of 92.23 percent on average by cleaning up the dataset and adding deep learning to our deep learning model. This is a significant improvement over the accuracy of the earlier models, which was just 85 percent. In addition, we have determined that, of all the algorithms that we have examined and evaluated, Deep Learning has the highest accuracy.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

REFERENCES

[1] Abu Nada, A. M., et al. (2020). "Age and Gender Prediction and Validation through Single User Images Using CNN." International Journal of Academic Engineering Research (IJAER) 4(8): 21-24.

[2] Abu Nada, A. M., et al. (2020). "Arabic Text Summarization Using AraBERT Model Using Extractive Text Summarization Approach." International Journal of Academic Information Systems Research (IJAISR) 4(8): 6-9.

[3] Abu-Jamie, T. N., et al. (2022). "Six Fruits Classification Using Deep Learning." International Journal of Academic Information Systems Research (IJAISR) 6(1): 1-8.

[4] Abu-Saqer, M. M., et al. (2020). "Type of Grapefruit Classification Using Deep Learning." International Journal of Academic Information Systems Research (IJAISR) 4(1): 1-5.

[5] Afana, M., et al. (2018). "Artificial Neural Network for Forecasting Car Mileage per Gallon in the City." International Journal of Advanced Science and Technology 124: 51-59.

[6] Aish, M. A., et al. (2022). "Classification of pepper Using Deep Learning." International Journal of Academic Engineering Research (IJAER) 6(1): 24-31.

[7] Al Barsh, Y. I., et al. (2020). "MPG Prediction Using Artificial Neural Network." International Journal of Academic Information Systems Research (IJAISR) 4(11): 7-16.

[8] Alajrami, E., et al. (2019). "Blood Donation Prediction using Artificial Neural Network." International Journal of Academic Engineering Research (IJAER) 3(10): 1-7.

[9] Alajrami, E., et al. (2020). "Handwritten Signature Verification using Deep Learning." International Journal of Academic Mult idisciplinary Research (IJAMR) 3(12): 39-44.

[10] Al-Araj, R. S. A., et al. (2020). "Classification of Animal Species Using Neural Network." International Journal of Academic Engineering Research (IJAER) 4(10): 23-31.

[11] Al-Araj, R. S. A., et al. (2020). "Classification of Animal Species Using Neural Network." International Journal of Academic Engineering Research (IJAER) 4(10): 23-31.

[12] Al-Atrash, Y. E., et al. (2020). "Modeling Cognitive Development of the Balance Scale Task Using ANN." International Journal of Academic Information Systems Research (IJAISR) 4(9): 74-81.

[13] Al-Atrash, Y. E., et al. (2020). "Modeling Cognitive Development of the Balance Scale Task Using ANN." International Journal of Academic

[14] Information Systems Research (IJAISR) 4(9): 74-81.

[15] Alfarra, A. H. et al. (2022). "Classification of Pineapple Using Deep Learning." International Journal of Academic Information Systems Research (IJAISR) 5(12): 37-41.

[16] Alghoul, A., et al. (2018). "Email Classification Using Artificial Neural Network." International Journal of Academic Engineering Research (IJAER) 2(11): 8-14.

[17] Alghoul, A., et al. (2018). "Email Classification Using Artificial Neural Network." International Journal of Academic Engineering Research (IJAER) 2(11): 8-14.

[18] Alkahlout, M. A. et al. (2022). "Classification of Fruits Using Deep Learning." International Journal of Academic Engineering Research (IJAER) 5(12): 56-63.

[19] Al-Kahlout, M. M., et al. (2020). "Neural Network Approach to Predict Forest Fires using Meteorological Data." International Journal of Academic Engineering Research (IJAER) 4(9): 68-72.

[20] Alkronz, E. S., et al. (2019). "Prediction of Whether Mushroom is Edible or Poisonous Using Back-propagation Neural Network." International Journal of Academic and Applied Research (IJAAR) 3(2): 1-8.

[21] Almadhoun, H. et al. (2021). "Classification of Alzheimer's Disease Using Traditional Classifiers with Pre-Trained CNN." International Journal of Academic Health and Medical Research (IJAHMR) 5(4):17-21.

[22] Almadhoun, H. R., et al. (2022). "Detection of Brain Tumor Using Deep Learning." International Journal of Academic Engineering Research (IJAER) 6(3).

[23] Al-Madhoun, O. S. E.-D., et al. (2020). "Low Birth Weight Prediction Using JNN." International Journal of Academic Health and Medical Research (IJAHMR) 4(11): 8-14.
[24] Al-Masawabe, M. M. et al. (2022). "Papaya maturity Classification Using Deep Convolutional Neural Networks." International Journal of Engineering and Information Systems (IJEAIS) 5(12): 60-67.
[25] Al-Massri, R., et al. (2018). "Classification Prediction of SBRCTs Cancers Using Artificial Neural Network." International Journal of Academic Engineering Research (IJAER) 2(11): 1-7.
[26] Al-Mobayed, A. A., et al. (2020). "Artificial Neural Network for Predicting Car Performance Using JNN." International Journal of Engineering and Information Systems (IJEAIS) 4(9): 139-145.
[27] Al-Mubayyed, O. M., et al. (2019). "Predicting Overall Car Performance Using Artificial Neural Network." International Journal o f Academic and Applied Research (IJAAR) 3(1): 1-5.
[28] Alsaqqa, A. H., et al. (2022). "Using Deep Learning to Classify Different types of Vitamin." International Journal of Academic Engineering Research (IJAER) 6(1): 1-6.
[29] Alshawwa, I. A., et al. (2020). "Analyzing Types of Cherry Using Deep Learning." International Journal of Academic Engineering Research (IJAER) 4(1): 1-5.
[30] Al-Shawwa, M., et al. (2018). "Predicting Temperature and Humidity in the Surrounding Environment Using Artificial Neural Network." International Journal of Academic Pedagogical Research (IJAPR) 2(9): 1-6.
[31] Ashqar, B. A., et al. (2019). "Plant Seedlings Classification Using Deep Learning." International Journal of Academic Information Systems Research (IJAISR) 3(1): 7-14.
[32] Bakr, M. A. H. A., et al. (2020). "Breast Cancer Prediction using JNN." International Journal of Academic Information Systems Research (IJAISR) 4(10): 1-8.
[33] Barhoom, A. M., et al. (2019). "Predicting Titanic Survivors using Artificial Neural Network." International Journal of Academic Engineering Research (IJAER) 3(9): 8-12.
[34] Barhoom, A. M., et al. (2022). "Diagnosis of Pneumonia Using Deep Learning." International Journal of Academic Engineering Research (IJAER) 6(2).
[35] Belbeisi, H. Z., et al. (2020). "Effect of Oxygen Consumption of Thylakoid Membranes (Chloroplasts) From Spinach after Inhibition Using JNN." International Journal of Academic Health and Medical Research (IJAHMR) 4(11): 1-7.
[36] Dalffa, M. A., et al. (2019). "Tic-Tac-Toe Learning Using Artificial Neural Networks." International Journal of Engineering and Information Systems (IJEAIS) 3(2): 9-19.
[37] Dawood, K. J., et al. (2020). "Artificial Neural Network for Mushroom Prediction." International Journal of Academic Information Systems Research (IJAISR) 4(10): 9-17.
[38] Dheir, I. M., et al. (2020). "Classifying Nuts Types Using Convolutional Neural Network." International Journal of Academic Information Systems Research (IJAISR) 3(12): 12-18.
[39] Dheir, I. M., et al. (2022). "Classification of Anomalies in Gastrointestinal Tract Using Deep Learning." International Journal of Academic Engineering Research (IJAER) 6(3).
[40] El-Habil, B. et al. (2022). "Cantaloupe Classification Using Deep Learning." International Journal of Academic Engineering Research (IJAER) 5(12): 7-17.
[41] El-Khatib, M. J., et al. (2019). "Glass Classification Using Artificial Neural Network." International Journal of Academic Pedagogical Research (IJAPR) 3(2): 25-31.
[42] El-Mahelawi, J. K., et al. (2020). "Tumor Classification Using Artificial Neural Networks." International Journal of Academic Engineering Research (IJAER) 4(11): 8-15.
[43] El-Mashharawi, H. Q., et al. (2020). "Grape Type Classification Using Deep Learning." International Journal of Academic Engineering Research (IJAER) 3(12): 41-45.
[44] Elsharif, A. A. E. F., et al. (2022). "Retina Diseases Diagnosis Using Deep Learning." International Journal of Academic Engineering Research (IJAER) 6(2).
[45] Elzamly, A., et al. (2015). "Classification of Software Risks with Discriminant Analysis Techniques in Software planning

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Development Process." International Journal of Advanced Science and Technology 81: 35-48.

[46] Elzamly, A., et al. (2015). "Predicting Software Analysis Process Risks Using Linear Stepwise Discriminant Analysis: Statistical Methods." Int. J. Adv. Inf. Sci. Technol 38(38): 108-115.

[47] Elzamly, A., et al. (2017). "Predicting Critical Cloud Computing Security Issues using Artificial Neural Network (ANNs) Algorithms in Banking Organizations." International Journal of Information Technology and Electrical Engineering 6(2): 40-45.

[48] Habib, N. S., et al. (2020). "Presence of Amphibian Species Prediction Using Features Obtained from GIS and Satellite Images." International Journal of Academic and Applied Research (IJAAR) 4(11): 13-22.

[49] Harz, H. H., et al. (2020). "Artificial Neural Network for Predicting Diabetes Using JNN." International Journal of Academic Engineering Research (IJAER) 4(10): 14-22.

[50] Hassanein, R. A. A., et al. (2020). "Artificial Neural Network for Predicting Workplace Absenteeism." International Journal o f Academic Engineering Research (IJAER) 4(9): 62-67.

[51] Heriz, H. H., et al. (2018). "English Alphabet Prediction Using Artificial Neural Networks." International Journal of Academic Pedagogical Research (IJAPR) 2(11): 8-14.

[52] Jaber, A. S., et al. (2020). "Evolving Efficient Classification Patterns in Lymphography Using EasyNN." International Journal of Academic Information Systems Research (IJAISR) 4(9): 66-73.

[53] Kashf, D. W. A., et al. (2018). "Predicting DNA Lung Cancer using Artificial Neural Network." International Journal of Academic Pedagogical Research (IJAPR) 2(10): 6-13.

[54] Khalil, A. J., et al. (2019). "Energy Efficiency Predicting using Artificial Neural Network." International Journal of Academic Pedagogical Research (IJAPR) 3(9): 1-8.

[55] Khalil, A. J., et al. (2022). "Diagnosis of Blood Cells Using Deep Learning." International Journal of Academic Engineering Research (IJAER) 6(2).

[56] Kweik, O. M. A., et al. (2020). "Artificial Neural Network for Lung Cancer Detection." International Journal of Academic Engineering Research (IJAER) 4(11): 1-7.

[57] Maghari, A. M., et al. (2020). "Books' Rating Prediction Using Just Neural Network." International Journal of Engineering and Information Systems (IJEAIS) 4(10): 17-22.

[58] Alnajjar, M. K., et al. (2022). "Heart Sounds Analysis and Classification for Cardiovascular Diseases Diagnosis using Deep Learning." International Journal of Academic Engineering Research (IJAER) 6(1): 7-23.

[59] Megdad, M. M., et al. (2022). "Fraudulent Financial Transactions Detection Using Machine Learning." International Journal of Academic Information Systems Research (IJAISR) 6(3).

[60] Mettleq, A. S. A., et al. (2020). "Mango Classification Using Deep Learning." International Journal of Academic Engineering Research (IJAER) 3(12): 22-29.

[61] Metwally, N. F., et al. (2018). "Diagnosis of Hepatitis Virus Using Artificial Neural Network." International Journal of Academic Pedagogical Research (IJAPR) 2(11): 1-7.

[62] Mohammed, G. R., et al. (2020). "Predicting the Age of Abalone from Physical Measurements Using Artificial Neural Network." International Journal of Academic and Applied Research (IJAAR) 4(11): 7-12.

[63] Musleh, M. M., et al. (2019). "Predicting Liver Patients using Artificial Neural Network." International Journal of Academic Information Systems Research (IJAISR) 3(10): 1-11.

[64] Obaid, T., et al. (2022). Factors Contributing to an Effective E-Government Adoption in Palestine. International Conference of Reliable Information and Communication Technology, Springer, Cham.

[65] Oriban, A. J. A., et al. (2020). "Antibiotic Susceptibility Prediction Using JNN." International Journal of Academic Informat ion Systems Research (IJAISR) 4(11): 1-6.

[66] Qwaider, S. R., et al. (2020). "Artificial Neural Network Prediction of the Academic Warning of Students in the Faculty of Engineering and Information Technology in Al-Azhar University-Gaza."

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

International Journal of Academic Information Systems Research (IJAISR) 4(8): 16-22.

[67] Sadek, R. M., et al. (2019). "Parkinson's Disease Prediction Using Artificial Neural Network." International Journal of Acade mic Health and Medical Research (IJAHMR) 3(1): 1-8.

[68] Salah, M., et al. (2018). "Predicting Medical Expenses Using Artificial Neural Network." International Journal of Engineering and Information Systems (IJEAIS) 2(20): 11-17.

[69] Saleh, A. et al. (2020). "Brain Tumor Classification Using Deep Learning." 2020 International Conference on Assistive and Rehabilitation Technologies (iCareTech). IEEE, 2020.

[70] Salman, F. M., et al. (2020). "COVID-19 Detection using Artificial Intelligence." International Journal of Academic Engineering Research (IJAER) 4(3): 18-25.

[71] Salman, F. M., et al. (2022). "Classification of Real and Fake Human Faces Using Deep Learning." International Journal of Academic Engineering Research (IJAER) 6(3).

[72] Samhan, L. F., et al. (2022). "Classification of Alzheimer's Disease Using Convolutional Neural Networks." International Journal of Academic Information Systems Research (IJAISR) 6(3).

[73] Samra, M. N. A., et al. (2020). "ANN Model for Predicting Protein Localization Sites in Cells." International Journal of Academic and Applied Research (IJAAR) 4(9): 43-50.

[74] Shawarib, M. Z. A., et al. (2020). "Breast Cancer Diagnosis and Survival Prediction Using JNN." International Journal of Engineering and Information Systems (IJEAIS) 4(10): 23-30.

[75] Zaqout, I., et al. (2015). "Predicting Student Performance Using Artificial Neural Network: in the Faculty of Engineering and Information Technology." International Journal of Hybrid Information Technology 8(2): 221-228.

[76] Aldammagh, Z., Abdeljawad, R., & Obaid, T. (2021). Predicting Mobile Banking Adoption: An Integration of TAM and TPB with Trust and Perceived Risk. Financial Internet Quarterly, 17(3), 35-46.

[77] Obaid, T. (2021). Predicting Mobile Banking Adoption: An Integration of TAM and TPB with Trust and Perceived Risk. Available at SSRN 3761669.

[78] Jouda, H., Abu Jarad, A., Obaid, T., Abu Mdallalah, S., &Awaja, A. (2020). Mobile Banking Adoption: Decomposed Theory of Planned Behavior with Perceived Trust. Available at SSRN 3660403.

[79] Obaid, T., Abdaljawad, R., &Mdallalahc, S. A. (2020). Factors Driving E-Learning Adoption In Palestine: An Integration of Technology Acceptance Model And IS Success Model. Available at SSRN 3686490.

[80] Obaid, T. F., &Eneizan, B. M. (2016). Transfer of training and post-training on job performance in Middle Eastern countries. Review of Public Administration and Management, 400(3786), 1-11.

[81] Obaid, T. F., Zainon, M. S., Eneizan, P. D. B. M., & Wahab, K. A. TRANSFER OF TRAINING AND POST-TRAINING ON JOB PERFORMANCE IN MIDDLE EASTERN COUNTRIES.

[82] MukteviSrivenkatesh(2020). Prediction of Cardiovascular Disease using Machine Learning Algorithms, International Journal of Engineering and Advanced Technology (IJEAT) Volume-9 Issue-3.

[83] Shah, D., Patel, S. & Bharti, S.K. Heart Disease Prediction using Machine Learning Techniques. SN COMPUT. SCI. 1, 345 (2020). https://doi.org/10.1007/s42979-020-00365-y

[84] Rohit Bharti et. al. (2021). Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning. Computational Intelligence and Neuroscience, https://doi.org/10.1155/2021/8387680.

[85] Abdul Saboor et. al. (2022). A Method for Improving Prediction of Human Heart Disease Using Machine Learning Algorithms. Mobile Information Systems. https://doi.org/10.1155/2022/1410169

[86] A. Nikam, S. Bhandari, A. Mhaske and S. Mantri, "Cardiovascular Disease Prediction Using Machine Learning Models," *2020 IEEE Pune Section International Conference (PuneCon)*, 2020, pp. 22-27, doi: 10.1109/PuneCon50868.2020.9362367.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[87] Aleem, A., Prateek, G., Kumar, N. (2022). Improving Heart Disease Prediction Using Feature Selection through Genetic Algorithm. In: Woungang, I., Dhurandher, S.K., Pattanaik, K.K., Verma, A., Verma, P. (eds) Advanced Network Technologies and Intelligent Computing. ANTIC 2021. Communications in Computer and Information Science, vol 1534. Springer, Cham. https://doi.org/10.1007/978-3-030-96040-7_57
[88] www.kaggle.com (last access on 15-3-2022)